

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
3 January 2003 (03.01.2003)

PCT

(10) International Publication Number  
**WO 03/000878 A2**

(51) International Patent Classification<sup>7</sup>: C12N 9/90

(21) International Application Number: PCT/EP02/06957

(22) International Filing Date: 24 June 2002 (24.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
01 115 225.3 22 June 2001 (22.06.2001) EP  
01 120 939.2 31 August 2001 (31.08.2001) EP

(71) Applicant (for DE only): **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant (for all designated States except DE): **F.HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstr. 124, CH-4070 Basel (CH).

(72) Inventors: **SCHOLZ, Christian**; Sindelsdorfer Strasse 35a, 82377 Penzberg (DE). **ANDRES, Herbert**; Kapellenwiese 39, 82377 Penzberg (DE). **FAATZ, Elke**; Steinkreuz 1, 82386 Huglfing (DE). **ENGEL, Alfred**; Kustermannstrasse 3a, 82327 Tutzing (DE). **SCHMITT,**

Urban; Waldstrasse 36, 82386 Oberhausen (DE). **BAZARSUREN, Ariuna**; Birkenstrasse 29, 82377 Penzberg (DE). **SCHAARSCHMIDT, Peter**; Birkenstrasse 4, 82390 Eberfing (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

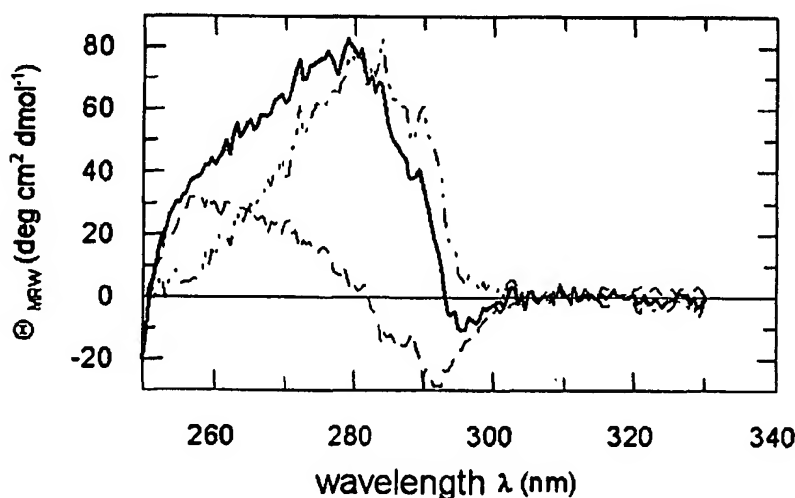
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: USE OF FKBP CHAPERONES AS EXPRESSION TOOL



(57) Abstract: The present invention relates to the cloning and expression of foreign protein or polypeptides in bacteria such as *Escherichia coli*. In particular, this invention relates to expression tools comprising a FKBP-type peptidyl prolyl isomerase selected from the group consisting of FkpA, SlyD, and trigger factor, methods of recombinant protein expression, the recombinant polypeptides thus obtained as well as to the use of such polypeptides.

### Use of FKBP chaperones as expression tool

The present invention relates to the cloning and expression of a heterologous protein or polypeptide in bacteria such as *Escherichia coli*. In particular, this invention relates to expression tools comprising a FKBP-type peptidyl prolyl isomerase selected from the group consisting of FkpA, SlyD, and trigger factor, methods of recombinant protein expression, 5 the recombinant polypeptides thus obtained as well as to the use of such polypeptides.

A large variety of expression systems has been described in the patent as well as in the scientific literature. However, despite the fact that fusion proteins have become a cornerstone of modern biology, obtaining the target protein in a soluble, biologically active form, as well as in high yield, continues to be a major challenge (Kapust, R. B. and Waugh, 10 D. S., Protein Sci 8 (1999) 1668-74).

Examples of fusion partners that have been touted as solubilizing agents include thioredoxin (TRX), glutathione S-transferase (GST), maltose-binding protein (MBP), Protein A, ubiquitin, and DsbA. Although widely recognized and potentially of great importance, this solubilizing effect remains poorly understood. It is not clear, for example, 15 what characteristics besides intrinsically high solubility epitomize an effective solubilizing agent. Are all soluble fusion partners equally proficient at this task, or are some consistently more effective than others? Similarly, it is not known whether the solubility of many different polypeptides can be improved by fusing them to a highly soluble partner or whether this approach is only effective in a small fraction of cases.

20 The state of the art relating to the most potent expression systems has recently been summarized by Kapust et al., *supra*. In their attempt to produce soluble fusion proteins comprising various target proteins they assessed three different and prominent candidate fusion partners. Maltose-binding protein (MBP), glutathione S-transferase (GST), and thioredoxin (TRX) have been tested for their ability to inhibit the aggregation of six diverse 25 proteins that normally accumulate in an insoluble form. All these candidate expression systems are known to the skilled artisan and described in detail elsewhere (e.g., EP 293 249 describes in detail the use of GST as an expression tool).

Remarkably, Kapust et al., *supra*, found that MBP is a far more effective solubilizing agent than the other two fusion partners also widely used in the art. Moreover, they 30 demonstrated that only in some cases fusion to MBP can promote the proper folding of the attached protein into its biologically active conformation.

It is especially critical that many aggregation-prone polypeptides may be rendered soluble by fusing them to an appropriate partner, but that some candidate fusion partners in a more or less unpredictable way are much better solubilizing agents than others.

While working on the recombinant expression of several retroviral surface glycoproteins (rsgps), we investigated the utility of many expression tools as known and recommended in the art, e.g. by Kapust et al., *supra*. However, we found that all the expression systems tested did suffer from one or several of the following shortcomings: low yield, fusion polypeptide difficult to handle, or insolubility of the fusion protein at physiological buffer conditions.

A great demand therefore exists to provide for alternative, efficient expression tools, which are especially appropriate for the recombinant expression of aggregation prone proteins, e.g. like the rsgps.

There is a wealth of patent literature relating to proteins which bind to the immunosuppressant FK-506, the so-called FK-506 binding proteins or FKBP.

These proteins have been extensively studied and commercial applications have been designed centering around the FK-506 binding activity of these proteins. For example, WO 93/25533 makes use of CTP:CMP-3-deoxy-D-manno-octulosonate cytidyl transferase (=CKS) as expression tool. A FKBP is inserted into a CKS-based expression vector downstream of the CKS gene. The fusion protein obtained is used to improve measurements of FK-506 and other immunosuppressants.

WO 00/28011 discloses materials and methods for regulation of biological events such as target gene transcription and growth, proliferation and differentiation of engineered cells.

WO 97/10253 relates to a high throughput assay for screening of compounds capable of binding to a fusion protein which consists of a target protein and an FK-506-binding protein. Disclosed is the use of a FKBP12-Src homology (SH2) fusion protein in an high throughput screening assay. The fusion protein is produced in soluble form in the bacterial periplasm and released by standard freeze-thaw treatment.

It was the task of the present invention to investigate whether it is possible to develop and provide efficient alternative expression systems which can be used for improved expression of a recombinant protein comprising a rsgp as a target protein and which at the same time are also appropriate for less critical target proteins.

To our surprise we have been able to identify certain modular members of the FKBP-type family of the peptidyl prolyl isomerase (PPI or PPIase) chaperones as very promising cloning tools. We found that an expression system based on a FKBP-type family of the chaperone selected from the group consisting of SlyD, FkpA, and trigger factor is ideal to  
5 express critical proteins like an rsgp and at the same time we could also demonstrate that these chaperones as well represent extremely promising cloning tools for less critical target proteins.

#### Summary of the invention

The present invention in a first embodiment relates to a recombinant DNA molecule,  
10 encoding a fusion protein, comprising at least one nucleotide sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP chaperone, characterized in that the FKBP chaperone is selected from the group consisting of FkpA, SlyD and trigger factor.

Preferred ways of designing such recombinant DNA molecules as well as their use as part of  
15 an expression vector, a host cell comprising such expression vector, and in the production of fusion polypeptide are also disclosed.

It has in addition been found that the recombinant fusion polypeptides themselves exhibit surprising and advantageous properties, e.g. with regard to solubilization, purification and handling. In a further embodiment the present invention relates to a recombinantly  
20 produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP chaperone selected from the group consisting of FkpA, SlyD and trigger factor and at least one polypeptide sequence corresponding to a target peptide.

A further embodiment relates to a recombinantly produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP chaperone selected from the  
25 group consisting of FkpA, SlyD and trigger factor, at least one polypeptide sequence corresponding to a target polypeptide, and at least one peptidic linker sequence of 10 - 100 amino acids.

Preferred recombinant fusion polypeptides are also disclosed as well as the use of such fusion polypeptides in various applications.

### Description of the Figures

#### Figure 1 UV spectrum of FkpA-gp41 at pH 2.5

UV-spectrum of the fusion polypeptide FkpA-gp41 after dialysis against 50 mM sodium phosphate, pH 2.5; 50 mM NaCl. Surprisingly, the two-domain construct remains completely soluble after removal of the solubilizing chaotropic agent GuHCl. There is no evidence for the existence of light-straying aggregates that would be expected to cause a baseline drift and significant apparent absorption at wavelengths beyond 300 nm.

#### Figure 2 Near UV CD spectrum of FkpA-gp41 at pH 2.5

The spectrum was recorded on a Jasco 720 spectropolarimeter in 20 mM sodium phosphate, pH 2.5; 50 mM NaCl at 20°C and was accumulated nine times to lower the noise. Protein concentration was 22.5 µM at a path length of 0.5 cm. The aromatic ellipticity shows the typical signature of gp41. At pH 2.5, FkpA is largely unstructured and does not contribute to the signal in the Near-UV-CD at all.

#### Figure 3 Far UV CD spectrum of FkpA-gp41 at pH 2.5

The spectrum was recorded on a Jasco 720 spectropolarimeter in 20 mM sodium phosphate pH 2.5; 50 mM NaCl at 20°C and was accumulated nine times to improve the signal-to-noise ratio. Protein concentration was 2.25 µM at a path-length of 0.2 cm. The minima at 220 and 208 nm point to a largely helical structure of gp41 in the context of the fusion protein. The spectral noise below 197 nm is due to the high amide absorption and does not report on any structural features of the fusion protein. Nevertheless, the typical helix-maximum at 193 nm can be guessed.

#### Figure 4 Near UV CD of FkpA-gp41 under physiological buffer conditions.

The spectrum was recorded on a Jasco 720 spectropolarimeter in 20 mM sodium phosphate, pH 7.4; 50 mM NaCl at 20°C and was accumulated nine times to lower the noise. Protein concentration was 15.5 µM at a path-length of 0.5 cm. Strikingly, the aromatic ellipticity of the covalently linked protein domains of gp41 and FkpA (continuous line) is made up additively from the contributions of native-like all-helical gp41 at pH 3.0 (lower dashed line) and the contributions of FkpA at pH 7.4 (upper dashed line). This indicates that the carrier FkpA and the target gp41 (i.e. two distinct functional folding

units) refold reversibly and quasi-independently when linked in a polypeptide fusion protein.

Figure 5 Far UV CD of FkpA-gp41 under physiological buffer conditions.

5 The spectrum was recorded on a Jasco 720 spectropolarimeter in 20 mM Sodium phosphate, pH 7.4; 50 mM NaCl at 20°C and accumulated nine times to improve the signal-to-noise ratio. Protein concentration was 1.55  $\mu$ M at a path-length of 0.2 cm. The strong signals at 222 nm and 208 nm, respectively, point to a largely helical structure of gp41 in the context of the fusion construct. The noise below 198 nm is due to the high protein absorption and does not reflect any secondary structural properties of FkpA-gp41.

10 Figure 6 The Near-UV-CD-spectra of scFkpA and scSlyD resemble each other

CD spectra were recorded on a Jasco-720 spectropolarimeter in 0.5 cm-cuvettes and averaged to improve the signal-to-noise-ratio. Buffer conditions were 50 mM sodium phosphate pH 7.8, 100 mM sodium chloride at 20 °C. Protein concentration was 45  $\mu$ M for both scFkpA (top line at 280 nm) and scSlyD (lower line at 280 nm), respectively. The  
15 structural similarity of both proteins is evidenced by the similar signature in the „fingerprint region“.

#### Detailed description

The present invention describes novel polypeptide expression systems. In a preferred embodiment it relates to a recombinant DNA molecule, encoding a fusion protein,  
20 comprising at least one nucleotide sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP chaperone, characterized in that the FKBP chaperone is selected from the group consisting of FkpA, SlyD and trigger factor.

As the skilled artisan will appreciate the term “at least one” is used to indicate that one or more nucleotide sequences coding for a target polypeptide, or for a FKBP chaperone,  
25 respectively, may be used in construction of a recombinant DNA molecule without departing from the scope of the present invention. Preferably the DNA construct will comprise one or two sequences coding for a target polypeptide, one being most preferred, and at the same time will contain at least one and at most four sequences coding for a chaperone, one or two being most preferred.

The term "recombinant DNA molecule" refers to a DNA molecule which is made by the combination of two otherwise separated segments of sequence accomplished by the artificial manipulation of isolated segments of polynucleotides by genetic engineering techniques or by chemical synthesis. In so doing one may join together polynucleotide  
5 segments of desired functions to generate a desired combination of functions.

Large amounts of the polynucleotides may be produced by replication in a suitable host cell. Natural or synthetic DNA fragments coding for proteins or fragments thereof will be incorporated into recombinant polynucleotide constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell.

- 10 The polynucleotides may also be produced by chemical synthesis, including, but not limited to, the phosphoramidite method described by Beaucage, S. L. and Caruthers, M. H., Tetrahedron Letters 22 (1981) 1859-1862 and the triester method according to Matteucci, M. D. and Caruthers, M. H., J. Am. Chem. Soc. 103 (1981) 3185-3191. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by  
15 synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

- A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods known in the art, the polynucleotide can be transcribed and/or  
20 translated to produce the polypeptide or a fragment thereof.

- A target polypeptide according to the present invention may be any polypeptide required in larger amounts and therefore difficult to isolate or purify from other non-recombinant sources. Examples of target proteins preferably produced by the present methods include mammalian gene products such as enzymes, cytokines, growth factors, hormones, vaccines,  
25 antibodies and the like. More particularly, preferred overexpressed gene products of the present invention include gene products such as erythropoietin, insulin, somatotropin, growth hormone releasing factor, platelet derived growth factor, epidermal growth factor, transforming growth factor  $\alpha$ , transforming growth factor  $\beta$ , epidermal growth factor, fibroblast growth factor, nerve growth factor, insulin-like growth factor I, insulin-like  
30 growth factor II, clotting Factor VIII, superoxide dismutase,  $\alpha$ -interferon,  $\gamma$ -interferon, interleukin-1, interleukin-2, interleukin-3, interleukin-4, interleukin-5, interleukin-6, granulocyte colony stimulating factor, multi-lineage colony stimulating activity, granulocyte-macrophage stimulating factor, macrophage colony stimulating factor, T cell

growth factor, lymphotoxin and the like. Preferred overexpressed gene products are human gene products. Moreover, the present methods can readily be adapted to enhance secretion of any overexpressed gene product which can be used as a vaccine. Overexpressed gene products which can be used as vaccines include any structural, membrane-associated, 5 membrane-bound or secreted gene product of a mammalian pathogen. Mammalian pathogens include viruses, bacteria, single-celled or multi-celled parasites which can infect or attack a mammal. For example, viral vaccines can include vaccines against viruses such as human immunodeficiency virus (HIV), vaccinia, poliovirus, adenovirus, influenza, hepatitis A, hepatitis B, dengue virus, Japanese B encephalitis, Varicella zoster, 10 cytomegalovirus, hepatitis A, rotavirus, as well as vaccines against viral diseases like measles, yellow fever, mumps, rabies, herpes, influenza, parainfluenza and the like. Bacterial vaccines can include vaccines against bacteria such as *Vibrio cholerae*, *Salmonella typhi*, *Bordetella pertussis*, *Streptococcus pneumoniae*, *Hemophilus influenza*, *Clostridium tetani*, *Corynebacterium diphtheriae*, *Mycobacterium leprae*, *R. rickettsii*, *Shigella*, *Neisseria* 15 *gonorrhoeae*, *Neisseria meningitidis*, *Coccidioides immitis*, *Borellia burgdorferi*, and the like.

Preferably, the target protein is a member of a group consisting of HIV-1 gp41, HIV-2 gp36, HTLV gp21, HIV-1 p17, SlyD, FkpA, and trigger factor.

A target polypeptide according to the present invention may also comprise sequences, e.g., diagnostically relevant epitopes, from several different proteins constructed to be expressed 20 as a single recombinant polypeptide.

The folding helpers termed peptidyl prolyl isomerases (PPIs or PPIases) are subdivided into three families, the parvulines (Schmid, F. X., Molecular chaperones in the life cycle of proteins (1998) 361-389, Eds. A. L. Fink and Y. Goto, Marcel Decker In., New York), Rahfeld, J. U., et al., FEBS Lett 352 (1994) 180-4) the cyclophilines (Fischer, G., et al., 25 Nature 337 (1989) 476-8, and the FKBP family (Lane, W. S., et al., J Protein Chem 10 (1991) 151-60). The FKBP family exhibits an interesting biochemical feature since its members have originally been identified by their ability to bind to macrolides, e.g., FK 506 and rapamycin (Kay, J. E., Biochem J 314 (1996) 361-85).

According to the present invention the preferred modular PPIases are FkpA (Ramm, K. 30 and Pluckthun, A., J Biol Chem 275 (2000) 17106-13), SlyD (Hottenrott, S., et al., J Biol Chem 272 (1997) 15697-701) and trigger factor (Scholz, C., et al., Embo J 16 (1997) 54-8), all members of the FKBP family. Most preferred are the chaperones FkpA and SlyD.



It is also well known and appreciated that it is not necessary to always use the complete sequence of a molecular chaperone. Functional fragments of chaperones (so-called modules) which still possess the required abilities and functions may also be used (cf. WO 98/13496).

- 5 For instance, FkpA is a periplasmic PPI that is synthesized as an inactive precursor molecule in the bacterial cytosol and translocated across the cytoplasmic membrane. The active form of FkpA (mature FkpA or periplasmic FkpA) lacks the signal sequence (amino acids 1 to 25) and thus comprises amino acids 26 to 270 of the precursor molecule. Relevant sequence information relating to FkpA can easily be obtained from public  
10 databases, e.g., from "SWISS-PROT" under accession number P 45523. The FkpA used as expression tool according to the present invention lacks the N-terminal signal sequence.

- A close relative of FkpA, namely SlyD, consists of a structured N-terminal domain responsible for catalytic and chaperone functions and of a largely unstructured C-terminus that is exceptionally rich in histidine and cysteine residues (Hottenrott, supra). We found  
15 that a C-terminally truncated variant of SlyD comprising amino acids 1-165 exerts exceptionally positive effects on the efficient expression of target proteins. Unlike in the wild-type SlyD, the danger of compromising disulfide shuffling is successfully circumvented in the truncated SlyD-variant (1-165) used. A recombinant DNA molecule comprising a truncated SlyD (1-165) represents a preferred embodiment of the present  
20 invention.

- In a preferred mode of designing a DNA construct according to the present invention no signal peptides are included. The expression systems according to the present invention have been found most advantageous when working as cytosolic expression system. This cytosolic expression results in the formation of inclusion bodies. Different from the  
25 pronounced and well-known problems usually associated with inclusion bodies, we now have found that not only an exceptionally high amount of protein is produced, but that the recombinant proteins according to the present invention are also easy to handle, e.g. easy to solubilize and to refold. In a preferred embodiment the present invention thus relates to a recombinant DNA molecule, encoding a fusion protein, comprising at least one nucleotide  
30 sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP chaperone, wherein the FKBP chaperone is selected from the group consisting of FkpA, SlyD and trigger factor further characterized in that the DNA construct lacks a signal peptide.

The term "lacks a signal peptide" must not be understood as an undue limitation. As the skilled artisan will readily appreciate either the construct may in fact lack the signal peptide sequence. As an alternative, however, the sequence may simply be modified to lack signal peptide function.

- 5 Variants of the above-discussed chaperones, bearing one or several amino acid substitutions or deletions, may also be used to obtain a recombinant DNA or a fusion polypeptide according to the present invention. The skilled artisan can easily assert whether such variants, e.g., fragments or mutants of chaperones or chaperones from alternative sources, are appropriate for a method of the invention by using the procedures as described  
10 in the Examples section.

- The term "recombinant" or "fusion polypeptide" as used in the present invention, refers to a polypeptide comprising at least one polypeptide domain corresponding to the FKBP-chaperone used as expression tool and at least one polypeptide domain corresponding to the target protein. Optionally such fusion protein may additionally comprise a linker  
15 polypeptide of 10 – 100 amino acid residues. As the skilled artisan will appreciate such linker polypeptide is designed as most appropriate for the intended application, especially in terms of length, flexibility, charge, and hydrophilicity.

- Preferably the DNA construct of the present invention encodes a fusion protein comprising a polypeptide linker in between the polypeptide sequence corresponding to the FKBP-chaperone and the polypeptide sequence corresponding to the target protein. Such DNA  
20 sequence coding for a linker in addition to e.g., provide for a proteolytic cleavage site, may also serve as a polylinker, i.e., it may provide multiple DNA restriction sites to facilitate fusion of the DNA fragments coding for a target protein and a chaperone domain.

- The present invention makes use of recombinant DNA technology in order to construct  
25 appropriate DNA molecules.

- In a further preferred embodiment the present invention relates a recombinant DNA molecule, encoding a fusion protein, comprising operably linked at least one nucleotide sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP chaperone, characterized in that the FKBP chaperone is  
30 selected from the group consisting of FkpA, SlyD and trigger factor.

Polynucleotide sequences are operably linked when they are placed into a functional relationship with another polynucleotide sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects transcription or expression of the coding sequence. Generally, operably linked means that the linked sequences are  
5 contiguous and, where necessary to join two protein coding regions, both contiguous and in reading frame. However, it is well known that certain genetic elements, such as enhancers, may be operably linked even at a distance, i.e., even if not contiguous.

As the skilled artisan will appreciate it is often advantageous to design a nucleotide sequence coding for a fusion protein such that one or a few, e.g., up to nine, amino acids  
10 are located in between the two polypeptide domains of said fusion protein. Fusion proteins thus constructed, as well as the DNA molecules encoding them obviously are also within the scope of the present invention.

DNA constructs prepared for introduction into a host typically comprise a replication system recognized by the host, including the intended DNA fragment encoding the desired  
15 target fusion peptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression systems (expression vectors) may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-  
20 binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences.

The appropriate promoter and other necessary vector sequences are selected so as to be functional in the host. Examples of workable combinations of cell lines and expression vectors include but are not limited to those described Sambrook, J., et al., in "Molecular  
25 Cloning: A Laboratory Manual" (1989) -, Eds. J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, or Ausubel, F., et al., in "Current protocols in molecular biology" (1987 and periodic updates), Eds. F. Ausubel, R. Brent and K. R.E., Wiley & Sons Verlag, New York; and Metzger, D., et al., Nature 334 (1988) 31-6. Many useful vectors for expression in bacteria, yeast, mammalian, insect, plant  
30 or other cells are known in the art and may be obtained from vendors including but not limited to Stratagene, New England Biolabs, Promega Biotech, and others. In addition, the construct may be joined to an amplifiable gene (e.g., DHFE) so that multiple copies of the gene may be obtained.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector, although such a marker gene may be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells expressing the marker gene will survive and/or grow under selective conditions. Typical selection genes include but are not limited to those encoding proteins that (a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, tetracycline, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are known in the art.

The vectors containing the polynucleotides of interest can be introduced into the host cell by any method known in the art. These methods vary depending upon the type of cellular host, including but not limited to transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, other substances, and infection by viruses. Large quantities of the polynucleotides and polypeptides of the present invention may be prepared by expressing the polynucleotides of the present invention in vectors or other expression vehicles in compatible host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* may also be used. Expression in *Escherichia coli* represents a preferred mode of carrying out the present invention.

Construction of a vector according to the present invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridization, using an appropriately labeled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In a preferred embodiment a recombinant DNA molecule according to the present invention comprises a single nucleotide sequence coding for a FKBP-chaperone selected

from the group consisting of FkpA, SlyD, and trigger factor and a single nucleotide sequence coding for a target polypeptide.

A fusion protein comprising two FKBP-chaperone domains and one target protein domain is also very advantageous. In a further preferred embodiment the recombinant DNA molecule according to the present invention comprises two sequences coding for a FKBP-chaperone and one sequence coding for a target polypeptide.

The DNA molecule may be designed to comprise both the DNA sequences coding for the FKBP-chaperone upstream to the target protein. Alternatively the two FKBP-domains may be arranged to sandwich the target protein. The construct comprising both FKBP-domains upstream to the target protein represents a preferred embodiment according to the present invention.

The DNA construct comprising two chaperone domains as well as a target polypeptide domain preferably also contains two linker peptides in between these domains. In order to allow for a systematic cloning the nucleotide sequences coding for these two linker peptide sequences preferably are different. This difference in nucleotide sequence must not necessarily result in a difference in the amino-acid sequence of the linker peptides. In yet a further preferred embodiment the amino acid sequences of the two linker peptides are identical. Such identical linker peptide sequences for example are advantageous if the fusion protein comprising two FKBP-chaperone domains as well as their target protein domain is to be used in an immunoassay.

In cases where it is desired to release one or all of the chaperones out of a fusion protein according to the present invention the linker peptide is constructed to comprise a proteolytic cleavage site. A recombinant DNA molecule encoding a fusion protein comprising at least one polypeptide sequence coding for a target polypeptide, upstream thereto at least one nucleotide sequence coding for a FKBP-chaperone selected from the group consisting of FkpA, SlyD, and trigger factor and additionally comprising a nucleic acid sequence coding for a peptidic linker comprising a proteolytic cleavage site, represents a further embodiment of this invention.

An expression vector comprising operably linked a recombinant DNA molecule according to the present invention, i.e., a recombinant DNA molecule encoding a fusion protein comprising at least one polynucleotide sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP-chaperone, wherein

the FKBP-chaperone is selected from FkpA, SlyD, and trigger factor, has proven to be very advantageous.

The expression vector comprising a recombinant DNA according to the present invention may be used to express the fusion protein in a cell free translation system or may be used to transform a host cell. In a preferred embodiment the present invention relates to a host cell transformed with an expression vector according to the present invention.

In a further preferred embodiment the present invention relates to a method of producing a fusion protein. Said method comprising the steps of culturing a host cell transformed with an expression vector according to the present invention, expression of that fusion protein in the respective host cell and purification of said fusion protein.

As discussed above the FKBP-chaperone domain of FkpA, SlyD, or trigger factor, respectively, is naturally or artificially constructed to yield a cytosolic fusion polypeptide expression. The fusion protein thus produced is obtained in form of inclusion bodies. Whereas in the art tremendous efforts are spent to obtain any desired recombinant protein or the fusion protein directly in a soluble form, we have found that the fusion protein according to the present invention is easily obtained in soluble form from inclusion bodies. In a further preferred embodiment the present invention therefore relates to a method of producing a fusion protein according to the steps described above, wherein said fusion protein is purified from inclusion bodies.

The purification of fusion protein from inclusion bodies is easily achieved and performed according to standard procedures known to the skilled artisan, like chaotropic solubilization and various ways of refolding.

Isolation and purification of the fusion protein starts from solubilizing buffer conditions, i.e. from a buffer wherein the inclusion bodies, i.e., the fusion protein, are/is solubilized. An appropriate buffer, which may be termed "non-physiological" or "solubilizing" buffer has to meet the requirement that both the target protein and the FKBP chaperone are not irreversibly denatured. Starting from such buffer conditions, the chaperone is in close proximity to the target protein, and a change of the buffer conditions from non-physiological to physiological conditions is possible without precipitation of the fusion protein.

An appropriate (non-physiological) buffer, i.e., a buffer wherein both the target protein which is essentially insoluble and the PPI-chaperone are soluble either makes use of high or low pH, or of a high chaotropic salt concentration or of a combination thereof. The solubilizing buffer preferably is a buffer with rather a high concentration of a chaotropic salt, e.g., 6.0 M guanidinium chloride at a pH of about 6. Upon renaturation both the target protein as well as the chaperone assume their native-like structure and the chaperone exerts its positive solubilizing effect..

In the context of this invention physiological buffer conditions are defined by a pH value between 5.0 and 8.5 and a total salt concentration below 500 mM, irrespective of other non-salt ingredients that optionally may be present in the buffer (e.g. sugars, alcohols, detergents) as long as such additives do not impair the solubility of the fusion protein comprising the target protein and the chaperone.

A variety of target proteins has been expressed in large amounts.

The expression system according to the present invention, for example, has been shown to work extremely well with biochemically rather different target proteins, e.g. SlyD, FkpA (proteins which are readily soluble), HIV-1 p17 (a protein which is difficult to express in high amounts using conventional expression systems), HTLV gp21 (a protein which tends to aggregate), and HIV-1 gp41, as well as HIV-2 gp36 (both proteins are extremely prone to aggregation and essentially insoluble under physiological buffer conditions). As can be easily gathered from Example 4 specifically relating to these proteins the efficient expression systems according to the present invention work and result in high levels of fusion protein produced. Similar positive findings have been made with a variety of other target proteins expressed as a fusion protein according to the present invention.

From the list of positive example it becomes readily obvious that the novel expression system as disclosed in the present invention, provide for extremely attractive universal expression systems.

The expression systems as disclosed herein also have been compared to standard expression systems making use of carrier proteins as recommended in the art, like MBP. It has been found that the novel systems with the target polypeptides tested are quite advantageous. The relative yield of fusion protein produced according to the present invention was at least as good and in the majority of cases even higher as compared to the relative yield using MBP-based expression. Efficacy of expression can be assessed both in terms of yield of

fusion protein, e.g., per g of *E. coli* cell mass or on a molar basis, comparing the concentrations of a target protein comprised in different fusion proteins.

The present invention in a preferred embodiment relates to a recombinantly produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP  
5 chaperone selected from the group consisting of FkpA, SlyD and trigger factor and at least one polypeptide sequence corresponding to a target peptide.

It has been found that the fusion proteins according to the present invention exhibit advantageous properties, thus e.g., facilitating production, handling and use of otherwise critical proteins. This becomes readily obvious from the description of the positive results  
10 obtained with a fusion protein comprising HIV-1 gp41. Whereas recombinantly produced gp41 itself is essentially insoluble, it is readily soluble if present as part of a fusion protein according to the present invention.

In general a protein is considered "essentially insoluble" if in a buffer consisting of 20 mM sodium phosphate pH 7.4, 150 mM NaCl it is soluble in a concentration of 50 nM or less. A  
15 fusion protein according to the present invention comprising a FKBP chaperone and a target protein is considered "soluble" if under physiological buffer conditions, e. g., in a buffer consisting of 20 mM sodium phosphate pH 7.4, 150 mM NaCl the target protein comprised in the PPI-chaperone complex is soluble in a concentration of 100 nM or more.

We found that the recombinantly produced fusion protein according to the present  
20 invention can be readily obtained from inclusion bodies in soluble form, even if the target protein is an aggregation prone protein like HIV-1 gp41. A striking feature of gp41 comprised in a recombinantly produced FkpA-gp41 is its exceptional solubility at physiological buffer conditions as compared to the "unchaperoned" gp41 ectodomain.

Moreover, it has been possible to demonstrate that the target protein comprised in a fusion  
25 protein according to the present invention readily can be obtained in a native-like structure. Such native-like structure, e.g., for HIV-1 gp41 has been confirmed by Near-UV-CD or by its immunoreactivity. Near-UV-CD analysis has shown the typical "gp41-signature" which is known to the skilled artisan.

The fusion protein according to the present invention also is very easy to handle, e.g., it is  
30 quite easy to renature such fusion protein. It is interesting that the "chaotropic material" (i.e. FkpA-gp41 in 6.0-7.0 M GuHCl) can be refolded in different ways, all resulting in a



thermodynamically stable and soluble native-like form. Refolding is achieved at high yields, both by dialysis and by rapid dilution, as well as by renaturing size exclusion chromatography or matrix-assisted refolding. These findings suggest that in this covalently linked form, the gp41-FkpA fusion polypeptide is a thermodynamically stable rather than a metastable protein.

Some of the FKBP-chaperones (e.g. FkpA) exert their chaperone function in form of oligomers, i.e., in a complex comprising two or more noncovalently associated FKBP polypeptides. We have surprisingly found that it is possible to design and produce such an active FKBP-dimer as a single fusion protein on one and the same polypeptide. We have termed these constructs single-chain PPIs, or single-chain FKBP. The single-chain PPI comprising two SlyD domains therefore is termed scSlyD and the single-chain PPI comprising two FkpA domains therefore is termed scFkpA. A single-chain peptidyl-prolyl-isomerase, i.e. a fusion protein comprising two PPI-domains represents a very advantageous and therefore preferred embodiment of the present invention. The sc-PPI according to the present invention may be a parvuline, a cyclophyline or a FKBP. The sc-PPIs selected from the FKBP family of chaperones are preferred. Most preferred are sc SlyD and Sc FkpA, respectively.

A recombinantly produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP chaperone selected from the group consisting of FkpA, SlyD and trigger factor, at least one polypeptide sequence corresponding to a target polypeptide, and at least one peptidic linker sequence of 10 - 100 amino acids represents a further preferred embodiment of the present invention.

As the skilled artisan will appreciate the peptidic linker may be constructed to contain the amino acids which are most appropriate for the required application. E.g., in case of a hydrophobic target protein the linker polypeptide preferably will contain an appropriate number of hydrophilic amino acids. The present invention specifically also relates to fusion proteins which comprise the target polypeptide and one, or two FKBP-chaperones or chaperone domains and an appropriate peptidic linker sequences between domains. For such applications where the target protein is required in free form a linker peptide or linker peptides are used, which contain an appropriate proteolytic cleavage site. Peptide sequences appropriate for proteolytic cleavage are well-known to the skilled artisan and comprise amongst others, e.g., Ile-Glu-Gly-Arg, cleaved at the carboxy side of the arginine residue by coagulation factor Xa, or Gly-Leu Pro-Arg-Gly-Ser, a thrombin cleavage site, etc..

- As mentioned above the fusion proteins according to the present invention can easily be obtained from inclusion bodies following a simple refolding scheme. They are readily soluble and target polypeptides comprised in such fusion proteins can easily be obtained in native-like confirmation. This is quite advantageous for polypeptides derived from an infectious organism because such native-like polypeptides are most advantageous in diagnostic as well as in therapeutic applications. In a preferred embodiment the fusion protein according to the present invention is further characterized in that a target protein is a polypeptide of interest as known from an infectious organism. Preferred infectious organisms according to the present invention are HIV, HTLV, and HCV.
- 10 From the scientific as well as from the patent literature it is well-known which peptide sequences contain diagnostically relevant epitopes. For the skilled artisan it is nowadays no problem to identify such relevant epitopes. In a further preferred embodiment the target protein corresponding to a polypeptide derived from an infectious organism will contain at least one diagnostically relevant epitope.
- 15 Due to their advantageous properties the recombinantly produced fusion proteins according to the present invention in further preferred embodiments are used for the immunization of laboratory animals, in the production of a vaccine or in an immunoassay, respectively.
- In case a therapeutic application of the novel fusion proteins is intended, preferably a composition comprising a recombinantly produced fusion protein according to the present invention and a pharmaceutically acceptable excipient will be formulated.
- 20
- The following examples, references, sequence listing and figures are provided to aid the understanding of the present invention, the true scope of which is set forth in the appended claims. It is understood that modifications can be made in the procedures set forth without departing from the spirit of the invention.
- 25

### Examples

**Example 1** Recombinant production of HIV-1 gp41 using an FkpA-based expression system

- 30 1.1 Construction of an expression plasmid comprising FkpA and gp41

- 18 -

Wild-type FkpA was cloned, expressed and purified according to Bothmann and Plückthun, *J Biol Chem* 275 (2000) 17106-17113 with some minor modifications. For storage, the protein solution was dialyzed against 20 mM NaH<sub>2</sub>PO<sub>4</sub>/NaOH (pH 6.0), 100 mM NaCl and concentrated to 26 mg/ml (1 mM).

- 5 For cytosolic expression, the FkpA-coding sequence of the above expression vector was modified to lack the sequence part coding for the signal peptide and to comprise instead only the coding region of mature FkpA.

In the first step, the restriction site *Bam*HI in the coding region of the mature *E. coli* FkpA was deleted using the QuikChange site-directed mutagenesis kit of Stratagene (La Jolla, CA; 10 USA) with the primers:

5'-gcgggtgttccgggtatcccaccgaattc-3' (SEQ ID NO: 1)

5'-gaattcgggtgggatacccggaacacccgc-3' (SEQ ID NO: 2)

The construct was named EcFkpA( $\Delta$  BamHI)[GGGS]<sub>3</sub>.

- 15 HIV-1 gp41 (535-681)-His<sub>6</sub> was cloned and expressed in a T7 promotor-based expression system. The gene fragment encoding amino acids 535-681 from HIV-1 envelope protein was amplified by PCR from the T7-based expression vector using the primers:

5'-cgggatccggtggcggttcaggcgggtggctctgggtggcggtacgctg-acggtacaggccag-3' (SEQ ID NO: 3)

5'-ccgctcgaggtaccacagccaattgttat-3' (SEQ ID NO: 4)

- 20 The fragment was inserted into EcFkpA( $\Delta$  BamHI)[GGGS]<sub>3</sub> using *Bam*HI and *Xho*I restriction sites.

The codons for a glycine-serine-rich linker [GGGS]<sub>3</sub> between FkpA and e-gp41 were inserted with reverse primer for cloning of FkpA and with forward primer for cloning of e-gp41.

- 25 The resulting construct was sequenced and found to encode the desired protein. Variants of this protein have also been generated by site-directed mutagenesis according to standard procedures. A variant of gp41 comprising four amino acid substitutions as compared to the

wild-type sequence is, e.g. encoded by the DNA-constructs of SEQ ID NO: 5 and 6, making use of FkpA or SlyD as expression system, respectively.

### 1.2 Purification of the FkpA-gp41 fusion protein from *E. coli* cells

*E. coli* BL21 cells harboring the expression plasmid were grown to a  $OD_{600}$  of 0.7, and  
5 cytosolic overexpression was induced by adding 1 mM of IPTG at a growth temperature of 37°C. Four hours after induction, the cells were harvested by centrifugation (20 min at 5000 g). The bacterial pellet was resuspended in 50 mM sodium phosphate pH 7.8, 6.0 M GuHCl (guanidinium chloride), 5 mM imidazole and stirred at room temperature (10 min) for complete lysis. After repeated centrifugation (Sorvall SS34, 20000 rpm, 4°C), the  
10 supernatant was filtered (0.8/0.2  $\mu$ m) and applied to a Ni-NTA-column (NTA: Nitrilotriacetate; Qiagen; Germantown, MD), pre-equilibrated in lysis buffer. Unspecifically bound proteins were removed in a washing step by applying 10 column volumes of lysis buffer. Finally, the bound target protein was eluted with 50 mM sodium phosphate, pH 2.5, 6.0 M GuHCl, and was collected in 4 ml fractions. The absorbance was  
15 recorded at 280 nm.

The resulting acidic and chaotropic solution may be stored at 4°C for further purification steps or *in vitro* refolding experiments.

Starting with this unfolded material, different refolding methods, such as dialysis, rapid dilution, renaturing size exclusion chromatography or matrix-assisted refolding can be  
20 used and carried out successfully, all of them leading to virtually the same native-like folded and soluble protein.

### 1.3 Renaturation by dialysis and rapid dilution

Material, solubilized as described above, is transferred into physiological buffer conditions by dialysis. The chosen cut-off value of the dialysis tubing was 4000 – 6000 Daltons.

25 To induce refolding of the ectodomain (the HIV-1 gp41 part of the fusion protein), GuHCl was removed from the eluted protein by dialysis against 50 mM sodium phosphate, pH 2.5, 50 mM NaCl (sodium chloride). It is well known that the isolated ectodomain is all-helical and forms tertiary contacts at this extreme pH. When analyzing recombinantly produced FkpA by means of near UV CD, it was found that FkpA is essentially unstructured under  
30 the same conditions. It is surprising that refolding of gp41-FkpA by dialysis results in a

readily soluble protein complex comprising the covalently linked gp41 and FkpA protein domains. The UV spectrum (Figure 1) lacks stray light, *i.e.*, apparent absorption beyond 300 nm. Stray light would be indicative of aggregates, thus the spectrum shown in Figure 1 implies that the re-folded material does not contain significant amounts of aggregates.

- 5 Circular dichroism spectroscopy (CD) is the method of choice to assess both secondary and tertiary structure in proteins. Ellipticity in the aromatic region (260-320 nm) reports on tertiary contacts within a protein (*i.e.*, the globular structure of a regularly folded protein), whereas ellipticity in the amide region reflects regular repetitive elements in the protein backbone, *i.e.*, secondary structure.
- 10 The near UV CD spectrum shown in Figure 2 provides compelling evidence that the ectodomain (in the context of the fusion protein) displays native-like tertiary contacts at pH 2.5. The spectrum of the covalently linked gp41/FkpA protein domains almost coincides with the spectrum of the isolated ectodomain under identical conditions (data not shown). The typical signature of gp41 was found: a maximum of ellipticity at 290 nm,
- 15 a characteristic shoulder at 285 nm and another maximum at 260 nm reflecting an optically active disulfide bridge. It is important to note that FkpA does not contribute to the near UV signal at all under the respective conditions. In fact, the aromatic ellipticity of FkpA at pH 2.5 virtually equals the baseline (data not shown).

- In agreement with the results from the near UV region, the far UV CD of the fusion
- 20 construct at pH 2.5 points to a largely structured gp41 molecule. The two maxima at 220 nm and 208 nm make up, and correspond to, the typical signature of an all-helical ectodomain (Figure 3). From the conditions indicated (50 mM sodium phosphate, pH 2.5, 50 mM NaCl), the FkpA-gp41 fusion polypeptide can easily be transferred to physiological buffer conditions by rapid dilution. In conclusion, both near and far UV CD underline that
- 25 native-like structured gp41 is available (in the context of the fusion protein also containing FkpA) in a very convenient fashion.

#### 1.4 Renaturation by size exclusion chromatography (SEC)

- Unfolded gp41-FkpA polypeptide (dissolved in 50 mM sodium phosphate, pH 7.8, 7.0 M GuHCl) was applied onto a Superdex 200 gel filtration column equilibrated with 20 mM
- 30 sodium phosphate, pH 7.4, 50 mM NaCl, 1 mM EDTA. FkpA-gp41 elutes essentially in three main fractions: as a high molecular associate, as an apparent hexamer species and as

an apparent trimer species. The apparent trimer fraction was concentrated and assessed for its tertiary structure in a near UV CD measurement (Figure 4).

The resulting graph is virtually an overlay curve to which both the carrier protein FkpA and the target protein gp41 contribute in a 1:1 ratio. Most fortunately, gp41 displays tertiary  
5 structure at neutral pH and is evidently solubilized by the covalently bound chaperone. In other words, the chaperone FkpA seems to accept the native-like structured ectodomain gp41 as a substrate and to solubilize this hard-to-fold protein at a neutral working pH. Thus, a crucial requirement for producing high amounts of soluble gp41 antigen for diagnostic purposes is fulfilled.

10 The far UV CD of FkpA-gp41 at pH 7.4 (Figure 5) confirms the near UV CD results in that it shows the additivity of the signal contributions of FkpA and gp41, respectively. As expected, the spectrum is dominated by the highly helical gp41 ectodomain (maximal ellipticity at 220 nm and 208 nm, respectively).

The data obtained with the covalently linked gp41/FkpA protein domains solubilized at pH  
15 7.4 under the conditions mentioned above indicate that FkpA and gp41 behave as independently folding units within the polypeptide construct.

#### Example 2 Use of a SlyD-based expression vector

The chaperone SlyD has been isolated by routine cloning procedures from *E. coli*. For recombinant expression a DNA construct has been prepared coding for amino acids 1 to  
20 165 of SlyD. An expression vector has been constructed comprising SlyD(1-165) as fusion partner and HIV-1 gp41 as target protein (cf.: SEQ ID NO: 6). The fusion protein was expressed and successfully purified as described for FkpA-gp41 above. Interestingly, we found that a native-like fusion polypeptide of the SlyD(1-165)-gp41 type can be obtained in a very convenient manner by dialysis of the chaotropic material (dissolved, e.g. in 7.0 M  
25 GuHCl) against 50 mM sodium phosphate pH 7.4, 150 mM NaCl at room temperature.

#### Example 3 Purification of scFkpA and scSlyD

The single-chain PPIases scSlyD (SEQ ID NO: 7) and scFkpA (SEQ ID NO: 8), respectively, were obtained from an *E. coli* overproducer according to virtually the same  
30 purification protocol as described in Example 1. In short: the induced cells were harvested, washed in PBS and lysed in 50 mM sodium phosphate pH 7.8, 100 mM sodium chloride,

7.0 M GuHCl at room temperature. The unfolded target proteins were bound to a Ni-NTA-column via their C-terminal hexa-His-tag and were refolded in 50 mM sodium phosphate pH 7.8, 100 mM sodium chloride. After this matrix-assisted refolding procedure, the proteins were eluted in an imidazole gradient and subjected to a gel  
5 filtration on a Superdex 200® column.

Alternatively, scSlyD and scFkpA may be dialysed after elution to remove residual concentrations of imidazole. Both proteins turn out to be highly soluble. ScSlyD, for example, does not tend to aggregation at concentrations up to 25 mg/ml. In order to elucidate the tertiary structure of the refolded scPPIases, we monitored CD-spectra in the  
10 Near-UV-region. The signatures of both scSlyD and scFkpA resemble each other and reflect the close relationship and thus structural homology of the two FKBP. Due to the low content in aromatic residues, the signal intensity of scSlyD (Fig. 6) is, however, significantly lower than the one of scFkpA.

#### Example 4 Improved expression of target proteins

15 The biochemically quite different target proteins HIV-1 gp41, HIV-2 gp36, HIV-1 p17 and HTLV gp21 have been expressed using the pET/BL21 expression system either without fusion partner (gp41, gp36, p17, gp21) or using same standard expression system but comprising a DNA-construct coding for a fusion protein according to the present invention (SlyD-gp41, FkpA-gp41, FkpA-p17, SlyD-gp36, FkpA-gp21). The efficiency of  
20 these systems has been compared in terms of yield of recombinant protein per *E. coli* cell mass [mg/g]. As becomes readily obvious from table 1, the novel expression systems lead to a significant improvement for all proteins tested.

Table 1:

Protein	Yield [mg protein/g E. coli cell mass]
gp41	~1-2
SlyD-gp41	~30
FkpA-gp41	~25
p17	~1
FkpA-p17	~15
gp36	~1-2
SlyD-gp36	~45
gp21	~4
FkpA-gp21	~30



List of References

- Sambrook, J., et al., in "Molecular Cloning: A Laboratory Manual" (1989) -, Eds. J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY
- 5 Beaucage, S. L. and Caruthers, M. H., Tetrahedron Letters 22 (1981) 1859-1862
- Ausubel, F., et al., in "Current protocols in molecular biology" (1987 and periodic updates), Eds. F. Ausubel, R. Brent and K. R.E., Wiley & Sons Verlag, New York
- Fischer, G., et al., Nature 337 (1989) 476-8
- Hottenrott, S., et al., J Biol Chem 272 (1997) 15697-701
- 10 Kapust, R. B. and Waugh, D. S., Protein Sci 8 (1999) 1668-74
- Kay, J. E., Biochem J 314 (1996) 361-85
- Lane, W. S., et al., J Protein Chem 10 (1991) 151-60
- Matteucci, M. D. and Caruthers, M. H., J. Am. Chem. Soc. 103 (1981) 3185-3191
- Metzger, D., et al., Nature 334 (1988) 31-6
- 15 Rahfeld, J. U., et al., FEBS Lett 352 (1994) 180-4
- Ramm, K. and Pluckthun, A., J Biol Chem 275 (2000) 17106-13
- Schmid, F. X., Molecular chaperones in the life cycle of proteins (1998) 361-389, Eds. A. L. Fink and Y. Goto, Marcel Decker In., New York
- Scholz, C., et al., Embo J 16 (1997) 54-8
- 20 EP 293 249
- WO 00/28011
- WO 93/25533
- WO 97/10253
- 25 WO 98/1349

## Patent Claims

1. A recombinant DNA molecule, encoding a fusion protein, comprising at least one nucleotide sequence coding for a target polypeptide and upstream thereto at least one nucleotide sequence coding for a FKBP chaperone, characterized in that the FKBP  
5 chaperone is selected from the group consisting of FkpA, SlyD and trigger factor.
2. The recombinant DNA molecule according to claim 1 further characterized in that it comprises at least one nucleotide sequence coding for a peptidic linker of 10 - 100 amino acids located in between said sequence coding for the target polypeptide and said sequence coding for the FKBP chaperone.
- 10 3. A recombinant DNA molecule according to claim 1 or 2, comprising one nucleotide sequence coding for said FKBP chaperone.
4. A recombinant DNA molecule according to claim 1 or 2, comprising two sequences coding for a FKBP chaperone.
- 15 5. The recombinant DNA molecule of claim 4 further characterized in that the two sequences coding for a FKBP chaperone are located upstream of the sequence coding for the target polypeptide.
6. The recombinant DNA molecule of claim 4 further characterized in that one sequence coding for a PPI chaperone is located upstream of the target polypeptide and the other sequence coding for a PPI chaperone is located downstream of the  
20 sequence coding for the target peptide.
7. The recombinant DNA molecule according to claim 4 to 6, further characterized in that, it comprises two nucleic acid sequences coding for a linker polypeptide of 10 - 100 amino acids.
8. The recombinant DNA molecule according to claim 7, wherein the two nucleic acid  
25 sequences coding for a linker of 10 - 100 amino acids are different.
9. The recombinant DNA molecule according to any of claims 2 to 8, wherein at least one of said linker sequences codes for a polypeptide linker comprising a proteolytic cleavage site.

- 26 -

10. An expression vector comprising operably linked a recombinant DNA molecule according to any of claims 1- 9.
11. A host cell transformed with an expression vector according to claim 10.
12. A method of producing a fusion protein said method comprising the steps of
  - 5 a. culturing host cells according to claim 11
  - b. expression of said fusion protein and
  - c. purification of said fusion protein.
13. A recombinantly produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP chaperone selected from the group consisting of  
10 FkpA, SlyD and trigger factor and at least one polypeptide sequence corresponding to a target peptide.
14. A recombinantly produced fusion protein comprising at least one polypeptide sequence corresponding to a FKBP chaperone selected from the group consisting of FkpA, SlyD and trigger factor, at least one polypeptide sequence corresponding to a  
15 target polypeptide, and at least one peptidic linker sequence of 10 - 100 amino acids.
15. The fusion protein according to claim 13 or 14, further characterized in that, it comprises one polypeptide sequence corresponding to said FKBP chaperone.
16. The fusion protein according to claim 13 or 14, further characterized in that, it comprises two polypeptide sequences corresponding to said FKBP chaperone
- 20 17. The fusion protein according to claim 16, further characterized in that, said two FKBP chaperones are located N-terminal with respect to the target polypeptide.
18. The fusion protein according to claim 16, further characterized in that, one of said two FKBP chaperones is located N-terminal and one of said FKBP chaperones is located C-terminal to the target polypeptide.

19. A recombinantly produced fusion protein comprising at least one target polypeptide, two sequences corresponding to FKBP chaperones selected from the group consisting of FkpA, SlyD and trigger factor and two peptidic linker sequences of 10 - 100 amino acids.
- 5 20. The fusion protein according to claim 19, wherein at least one of said peptidic linker sequences comprises a proteolytic cleavage site.
21. The fusion protein according to any of claims 13 - 20, wherein said target protein comprises a polypeptide from an infectious organism.
- 10 22. The fusion protein according to claims 21, further characterized in that said polypeptide comprises at least one diagnostically relevant epitope of an infectious organism.
23. Use of a recombinantly produced fusion protein according to any of claims 13 - 22, for immunization of laboratory animals.
- 15 24. Use of a recombinantly produced fusion protein according to any of claims 13 - 22, in the production of a vaccine.
25. Use of a recombinantly produced fusion protein according to any of claims 13 - 22, in an immunoassay.
26. A composition comprising a recombinantly produced fusion protein according to any of claims 13 - 22, and a pharmaceutically acceptable excipient.

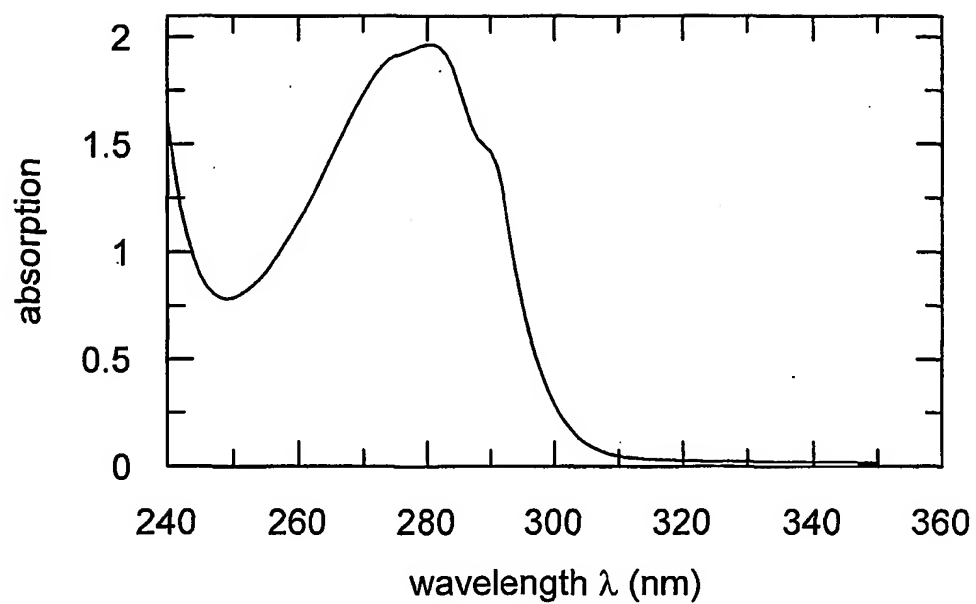
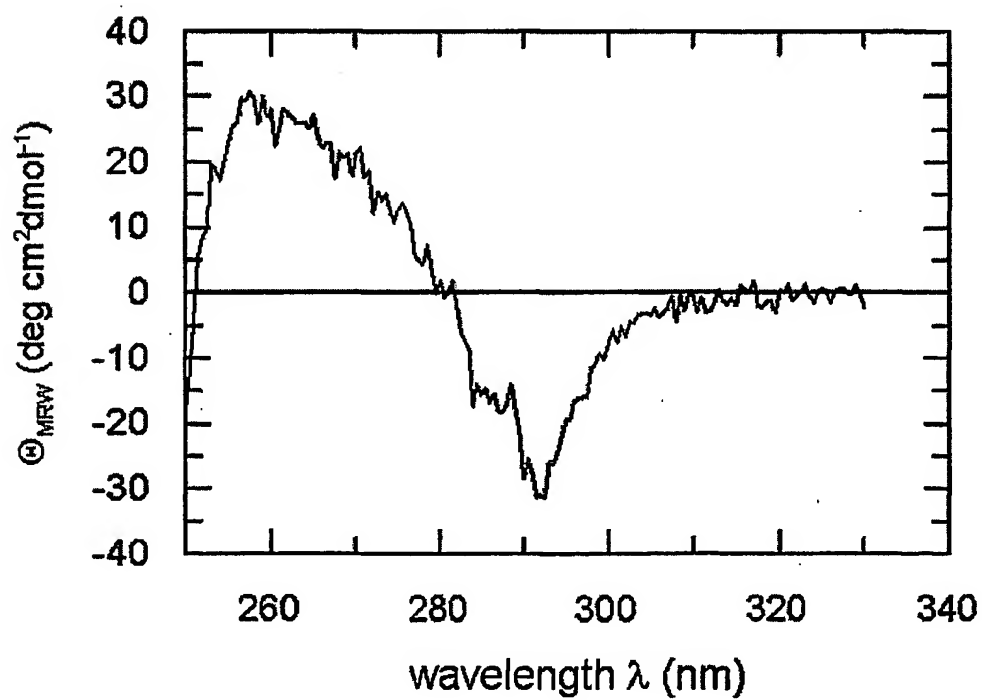
Figure 1:Figure 2

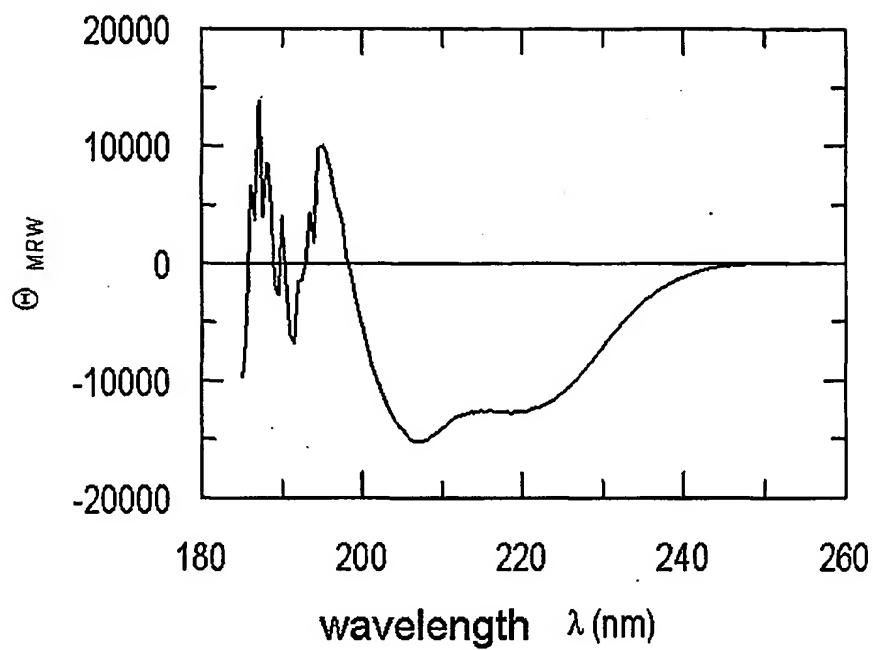
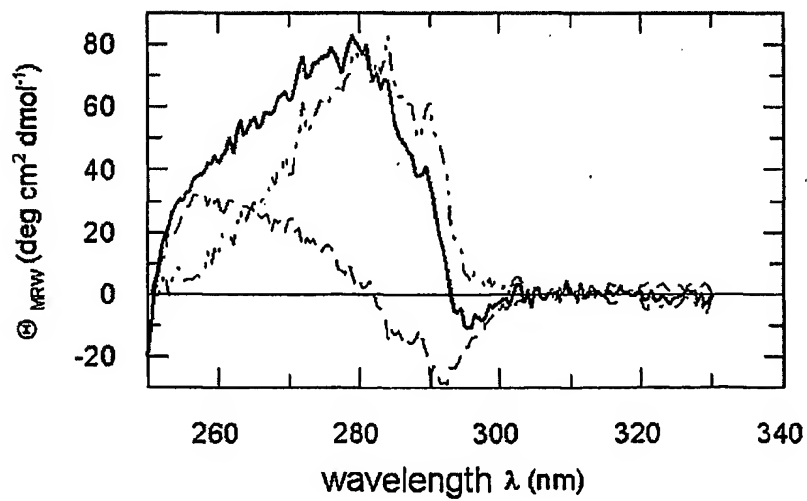
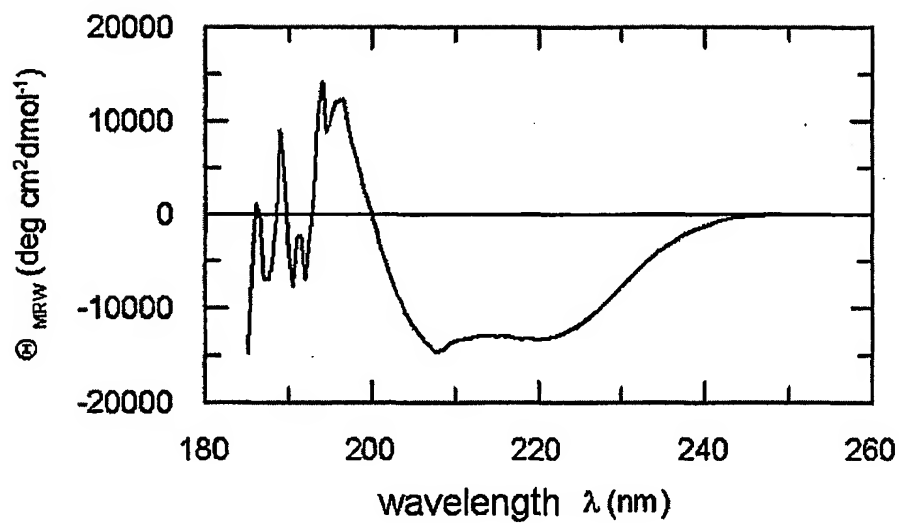
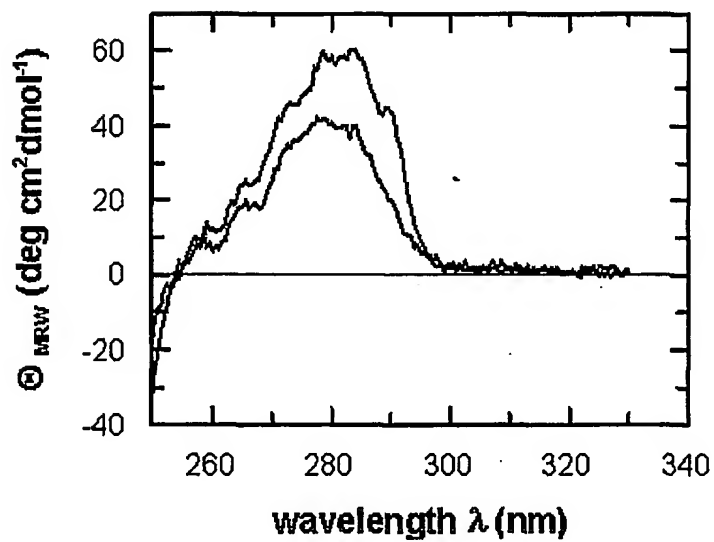
Figure 3Figure 4

Figure 5Figure 6

- 1 -

## Sequenzprotokoll

<110> Roche Diagnostics GmbH  
F. Hoffmann-La Roche AG

5 <120> Use of FKBP chaperones as expression tool

<130> 21306WO

10 <140>  
<141>

<150> EP01115225.3  
<151> 2001-06-22

15 <150> EP01120939.2  
<151> 2001-08-31

<160> 8

20 <170> PatentIn Ver. 2.1

<210> 1  
<211> 29

25 <212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:primer 1

30 <400> 1  
gcgggtgttc cgggtatccc accgaattc 29

35 <210> 2  
<211> 29  
<212> DNA  
<213> Artificial Sequence

40 <220>  
<223> Description of Artificial Sequence:primer 2

<400> 2  
gaattcgggtg ggatacccg g aacacccgc 29

45 <210> 3  
<211> 61  
<212> DNA

50 <213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:primer 3

55 <400> 3



- 2 -

cgggatccgg tggcggttca ggcggtggct ctggtggcgg tacgctgacg gtacaggcca 60  
g 61

5 <210> 4  
<211> 30  
<212> DNA  
<213> Artificial Sequence

10 <220>  
<223> Description of Artificial Sequence:primer 4

<400> 4  
ccgctcgagg taccacagcc aatttggtat 30

15

<210> 5  
<211> 1269  
<212> DNA  
20 <213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:coding for a  
FkpA-gp41 fusion protein

25

<400> 5  
atggctgaag ctgcaaaacc tgctacaact gctgacagca aagcagcgtt caaaaatgac 60  
gatcagaaat cagcttatgc actgggtgct tcgctgggtc gttacatgga aaactctctt 120  
aaagaacaag aaaaactggg catcaaactg gataaagatc agctgatcgc tgggtgttcag 180  
30 gatgcatttg ctgataagag caaactctcc gaccaagaga tcgaacagac tctgcaagca 240  
ttcgaagctc gctggaagtc ttctgctcag gcgaagatgg aaaaagacgc ggctgataac 300  
gaagcaaaag gtaaagagta ccgcgagaaa tttgccaaag agaaagggtg gaaaacctct 360  
tcaactggtc tggtttatca ggtagtagaa gccggtaaag gcgaagcacc gaaagacagc 420  
gatactgttg tagtgaacta caaaggtagc ctgatcgacg gtaaagagtt cgacaactct 480  
35 tacacccgtg gtgaaccgct ctctttccgt ctggacgggtg ttatcccggg ttggacagaa 540  
ggtctgaaga acatcaagaa aggcggtaag atcaaactgg ttattccacc agaactggct 600  
tacggcaaag cgggtgttcc gggtatocca ccgaattcta ccctgggtgt tgacgtagag 660  
ctgctggatg tgaaccagc gccgaaggct gatgcaaagc cggaagctga tgcgaaagcc 720  
gcagattctg ctaaaaaagg tggcggttcc ggcggtggct ctggtggcgg atccggtggc 780  
40 ggttccggcg gtggctctgg tggcggtacg ctgacggtac aggccagaca attattgtct 840  
ggtatagtgc agcagcagaa caatgagctg agggctattg aggcgcaaca gcatctggag 900  
caactcacag tctggggcac caagcagctc caggcaagag aactggctgt ggaaagatac 960  
ctaaaggatc aacagctcct ggggatttgg ggttgctctg gaaaactcat ttgcaccact  
1020  
45 gctgtgcctt ggaatgctag ttggagtaat aaatctctgg aacagatttg gaataacatg  
1080  
acctggatgg agtgggacag agaaattaac aattacacaa gcttaataca ttccttaatt  
1140  
gaagaatcgc aaaaccagca agaaaagaat gaacaagaat tattggaatt agataaatgg  
50 1200  
gcaagtttgt ggaattgggt taacataaca aattggctgt ggtacctcga gcaccaccac  
1260  
caccaccac  
1269  
55

- 3 -

<210> 6  
 <211> 1026  
 <212> DNA  
 <213> Artificial Sequence

5

<220>  
 <223> Description of Artificial Sequence: coding for a  
 SlyD-gp41 fusion protein

10 <400> 6

atgaaagtag	caaaagacct	ggtggtcagc	ctggcctatc	aggtacgtac	agaagacggt	60
gtgttggttg	atgagtcctc	ggtgagtgcg	ccgctggact	acctgcatgg	tcacggttcc	120
ctgatctctg	gcctggaaac	ggcgctggaa	ggcatgaag	ttggcgacaa	atttgatgtc	180
gctgttggcg	cgaacgacgc	ttacggtcag	tacgacgaaa	acctggtgca	acgtgttcct	240
15 aaagacgtat	ttatgggcgt	tgatgaactg	caggtaggta	tgcgtttcct	ggctgaaacc	300
gaccagggtc	cggtaccggt	tgaaatcact	gcggttgaag	acgatcacgt	cgtggttgat	360
ggtaaccaca	tgctggccgg	tcagaacctg	aaattcaacg	ttgaagttgt	ggcgattcgc	420
gaagcgactg	aagaagaact	ggctcatggt	cacgttcacg	gcgcgcacga	tcaccaccac	480
gatcacgacc	acgacgggtg	cggttcggcg	ggtggctctg	gtggcggatc	cggtggcggt	540
20 tccggcggtg	gctctggtgg	cggtacgctg	acggtacagg	ccagacaatt	attgtctggt	600
atagtgcagc	agcagaacaa	tgagctgagg	gctattgagg	cgcaacagca	tctggagcaa	660
ctcacagtct	ggggcaccaa	gcagctccag	gcaagagaac	tggtctgtga	aagataccta	720
aaggatcaac	agctcctggg	gatttggggg	tgctctggaa	aactcatttg	caccactgct	780
gtgccttgga	atgctagtgt	gagtaataaa	tctctggaac	agatttggaa	taacatgacc	840
25 tggatggagt	gggacagaga	aattaacaat	tacacaagct	taatacattc	cttaattgaa	900
gaatcgcaaa	accagcaaga	aaagaatgaa	caagaattat	tggaattaga	taaattgggca	960
agtttgtgga	attggtttaa	cataacaaat	tggtctgtgt	acctcgagca	ccaccaccac	

1020

caccac

30 1026

<210> 7  
 <211> 367  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: single-chain  
 SlyD

40

<400> 7

Met	Lys	Val	Ala	Lys	Asp	Leu	Val	Val	Ser	Leu	Ala	Tyr	Gln	Val	Arg	
1				5					10					15		
45	Thr	Glu	Asp	Gly	Val	Leu	Val	Asp	Glu	Ser	Pro	Val	Ser	Ala	Pro	Leu
				20					25					30		
	Asp	Tyr	Leu	His	Gly	His	Gly	Ser	Leu	Ile	Ser	Gly	Leu	Glu	Thr	Ala
50			35					40					45			
	Leu	Glu	Gly	His	Glu	Val	Gly	Asp	Lys	Phe	Asp	Val	Ala	Val	Gly	Ala
		50					55					60				
55	Asn	Asp	Ala	Tyr	Gly	Gln	Tyr	Asp	Glu	Asn	Leu	Val	Gln	Arg	Val	Pro
	65					70					75				80	

- 4 -

Lys Asp Val Phe Met Gly Val Asp Glu Leu Gln Val Gly Met Arg Phe  
                             85                            90                            95  
 5 Leu Ala Glu Thr Asp Gln Gly Pro Val Pro Val Glu Ile Thr Ala Val  
                             100                            105                            110  
 Glu Asp Asp His Val Val Val Asp Gly Asn His Met Leu Ala Gly Gln  
                             115                            120                            125  
 10 Asn Leu Lys Phe Asn Val Glu Val Val Ala Ile Arg Glu Ala Thr Glu  
                             130                            135                            140  
 Glu Glu Leu Ala His Gly His Val His Gly Ala His Asp His His His  
 15 145                            150                            155                            160  
 Asp His Asp His Asp Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly  
                             165                            170                            175  
 20 Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Lys Val Ala Lys  
                             180                            185                            190  
 Asp Leu Val Val Ser Leu Ala Tyr Gln Val Arg Thr Glu Asp Gly Val  
                             195                            200                            205  
 25 Leu Val Asp Glu Ser Pro Val Ser Ala Pro Leu Asp Tyr Leu His Gly  
                             210                            215                            220  
 His Gly Ser Leu Ile Ser Gly Leu Glu Thr Ala Leu Glu Gly His Glu  
 30 225                            230                            235                            240  
 Val Gly Asp Lys Phe Asp Val Ala Val Gly Ala Asn Asp Ala Tyr Gly  
                             245                            250                            255  
 35 Gln Tyr Asp Glu Asn Leu Val Gln Arg Val Pro Lys Asp Val Phe Met  
                             260                            265                            270  
 Gly Val Asp Glu Leu Gln Val Gly Met Arg Phe Leu Ala Glu Thr Asp  
                             275                            280                            285  
 40 Gln Gly Pro Val Pro Val Glu Ile Thr Ala Val Glu Asp Asp His Val  
                             290                            295                            300  
 Val Val Asp Gly Asn His Met Leu Ala Gly Gln Asn Leu Lys Phe Asn  
 45 305                            310                            315                            320  
 Val Glu Val Val Ala Ile Arg Glu Ala Thr Glu Glu Glu Leu Ala His  
                             325                            330                            335  
 50 Gly His Val His Gly Ala His Asp His His His Asp His Asp His Asp  
                             340                            345                            350  
 Gly Gly Gly Ser Gly Gly Gly Leu Glu His His His His His His  
                             355                            360                            365  
 55

- 5 -

<210> 8  
 <211> 537  
 <212> PRT  
 <213> Artificial Sequence  
 5  
 <220>  
 <223> Description of Artificial Sequence:single-chain  
 FkpA  
 10 <400> 8  
 Met Ala Glu Ala Ala Lys Pro Ala Thr Thr Ala Asp Ser Lys Ala Ala  
 1 5 10 15  
 Phe Lys Asn Asp Asp Gln Lys Ser Ala Tyr Ala Leu Gly Ala Ser Leu  
 15 20 25 30  
 Gly Arg Tyr Met Glu Asn Ser Leu Lys Glu Gln Glu Lys Leu Gly Ile  
 35 40 45  
 20 Lys Leu Asp Lys Asp Gln Leu Ile Ala Gly Val Gln Asp Ala Phe Ala  
 50 55 60  
 Asp Lys Ser Lys Leu Ser Asp Gln Glu Ile Glu Gln Thr Leu Gln Ala  
 65 70 75 80  
 25 Phe Glu Ala Arg Val Lys Ser Ser Ala Gln Ala Lys Met Glu Lys Asp  
 85 90 95  
 Ala Ala Asp Asn Glu Ala Lys Gly Lys Glu Tyr Arg Glu Lys Phe Ala  
 30 100 105 110  
 Lys Glu Lys Gly Val Lys Thr Ser Ser Thr Gly Leu Val Tyr Gln Val  
 115 120 125  
 35 Val Glu Ala Gly Lys Gly Glu Ala Pro Lys Asp Ser Asp Thr Val Val  
 130 135 140  
 Val Asn Tyr Lys Gly Thr Leu Ile Asp Gly Lys Glu Phe Asp Asn Ser  
 145 150 155 160  
 40 Tyr Thr Arg Gly Glu Pro Leu Ser Phe Arg Leu Asp Gly Val Ile Pro  
 165 170 175  
 Gly Trp Thr Glu Gly Leu Lys Asn Ile Lys Lys Gly Gly Lys Ile Lys  
 45 180 185 190  
 Leu Val Ile Pro Pro Glu Leu Ala Tyr Gly Lys Ala Gly Val Pro Gly  
 195 200 205  
 50 Ile Pro Pro Asn Ser Thr Leu Val Phe Asp Val Glu Leu Leu Asp Val  
 210 215 220  
 Lys Pro Ala Pro Lys Ala Asp Ala Lys Pro Glu Ala Asp Ala Lys Ala  
 225 230 235 240  
 55 Ala Asp Ser Ala Lys Lys Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly

- 6 -

		245		250		255
	Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly	260	265	270		
5	Gly Ser Gly Gly Gly Ala Glu Ala Ala Lys Pro Ala Thr Thr Ala Asp	275	280	285		
10	Ser Lys Ala Ala Phe Lys Asn Asp Asp Gln Lys Ser Ala Tyr Ala Leu	290	295	300		
	Gly Ala Ser Leu Gly Arg Tyr Met Glu Asn Ser Leu Lys Glu Gln Glu	305	310	315		320
15	Lys Leu Gly Ile Lys Leu Asp Lys Asp Gln Leu Ile Ala Gly Val Gln	325	330	335		
	Asp Ala Phe Ala Asp Lys Ser Lys Leu Ser Asp Gln Glu Ile Glu Gln	340	345	350		
20	Thr Leu Gln Ala Phe Glu Ala Arg Val Lys Ser Ser Ala Gln Ala Lys	355	360	365		
	Met Glu Lys Asp Ala Ala Asp Asn Glu Ala Lys Gly Lys Glu Tyr Arg	370	375	380		
25	Glu Lys Phe Ala Lys Glu Lys Gly Val Lys Thr Ser Ser Thr Gly Leu	385	390	395		400
30	Val Tyr Gln Val Val Glu Ala Gly Lys Gly Glu Ala Pro Lys Asp Ser	405	410	415		
	Asp Thr Val Val Val Asn Tyr Lys Gly Thr Leu Ile Asp Gly Lys Glu	420	425	430		
35	Phe Asp Asn Ser Tyr Thr Arg Gly Glu Pro Leu Ser Phe Arg Leu Asp	435	440	445		
	Gly Val Ile Pro Gly Trp Thr Glu Gly Leu Lys Asn Ile Lys Lys Gly	450	455	460		
40	Gly Lys Ile Lys Leu Val Ile Pro Pro Glu Leu Ala Tyr Gly Lys Ala	465	470	475		480
45	Gly Val Pro Gly Ile Pro Pro Asn Ser Thr Leu Val Phe Asp Val Glu	485	490	495		
	Leu Leu Asp Val Lys Pro Ala Pro Lys Ala Asp Ala Lys Pro Glu Ala	500	505	510		
50	Asp Ala Lys Ala Ala Asp Ser Ala Lys Lys Gly Gly Gly Ser Gly Gly	515	520	525		
55	Gly Leu Glu His His His His His His	530	535			